European Journal of Operational Research 298 (2022) 1016–1031

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

Public health interventions in the face of pandemics: Network structure, social distancing, and heterogeneity

Mohammad Ghaderi [a,b,c]

[a] Department of Economics and Business, Pompeu Fabra University, Barcelona, Spain
[b] Barcelona Graduate School of Economics, Barcelona, Spain
[c] UPF Barcelona School of Management, Spain

ARTICLE INFO

Article history:
Received 4 August 2020
Accepted 11 August 2021
Available online 25 August 2021

Keywords:
Decision support systems
Contagion
Agent-based models
Random networks
Stochastic simulation

ABSTRACT

Complexity, resulting from interactions among many components, is a characterizing feature of healthcare systems and related decisions. It scales up in the face of pandemics that give rise to multiple sources of uncertainty and where various contextual factors interact with each other and with policy parameters that combine to yield outcome distributions. This paper proposes a unified agent-based modeling framework to derive qualitative insights that assist and inform policy decisions related to pandemics. The general framework comprises a contagion model that explicates exogenous policy-relevant variables, as well as their links with features of the environment in which the policy decisions will be implemented. Furthermore, the framework identifies sources of uncertainty at different system layers. The characterization of the macro level, for example, as manifested in the network structure, encompasses two constitutive factors. These two factors, in turn, capture much of the stochasticity that results from the network's inherent randomness. By synthesizing the model components further into a broader agent-based model, the current framework also accounts for heterogeneous micro-level attributes that collectively yield macro-level outcomes. Several stylized examples help establish insights into the overall tendency of complex systems to produce multidimensional outputs. A comprehensive, controlled, computational experiment offers further evidence across a range of scenarios and various policy conditions.

1. Introduction

The novel coronavirus (COVID-19) outbreak represents a public health emergency of international concern, challenging government, public, and private sectors with multifaceted impacts. In particular, the rapid spread of the disease has overwhelmed public health systems. The various intervention strategies adopted by governments worldwide mostly rely on containment and non-pharmaceutical measures, such as household quarantine, workplace closures, and mobility restrictions. But beyond these health considerations, the pandemic and related containment strategies have imposed severe economic damages with potentially long-lasting impacts, including losses of employment, disruptions of global supply chains, adverse supply shocks, and demand shortages (Fernandes, 2020; Fornaro & Wolf, 2020; McKibbin & Fernando, 2020). The COVID-19 pandemic also has prompted substantial global attention and media coverage, yet it joins many other disease outbreaks in recent decades, such as Ebola in 2014 and influenza in 2009. The situation forewarns similar incidents, especially considering the rapid, global increases in international travel and trade, facilitating the rapid circulation of goods, people, and infectious diseases. Developing effective intervention strategies to mitigate the impacts of the persistent COVID-19 and future pandemics thus is a critical and global priority.

Designing an effective response plan for confronting pandemics represents a highly complex task, for three main reasons. First, the substantial uncertainty involved, combined with a lack of sufficient data, prevent reliable predictions about the effectiveness of alternative intervention strategies and their multidimensional consequences, pertaining to healthcare, economic outcomes, and social benefits. Such complexity is particularly challenging when decision makers must compromise and prioritize various and conflicting decision criteria.

Second, assessing the relative efficacy of alternative policy choices is a cumbersome task. For example, self-isolation methods might limit the spread of viruses, but they simultaneously create substantial economic and mental health concerns. For policymak-

E-mail address: mohammad.ghaderi@upf.edu

https://doi.org/10.1016/j.ejor.2021.08.015

ing purposes, qualitative insights about the relative effectiveness of the multiple possible alternative intervention methods are critical. Such assessments remain difficult though, in that they must account for a vast range of possibilities, such as healthcare capacity expansions, health service quality improvements, expanded testing capacity, public area decontamination, or any combination of thereof, while also taking budget constraints into account. Well-informed decisions demand insights into how the different intervention methods compare but also how they might interact within a unified response plan.

Third, the surrounding environment can largely determine the efficacy of an intervention method, which in turn is influenced by the policy adopted. Thus, the mechanism by which policy parameters affect an outcome distribution can be shaped powerfully by the dynamic characteristics of the environment. For example, the frequency of contacts by people who adopt heterogeneous social distancing behavior can determine the infection transmission rate. A population's overall health status and the capacity available in a healthcare system, combined with the total number of hospitalized individuals, in turn, determined by the infection transmission rate and policy parameters, also determine the load imposed on the healthcare system. This burden then informs health service quality, which ultimately affects the mortality rate. Even with a particular population, heterogeneity in people's views and behaviors influence policy outcomes, such that for people with insecure housing, sheltering at home policies are ineffective. In this case, people's self-isolating behavior differs across any particular population, based on individual capacity to adopt it (Chiou & Tucker, 2020). Effective policy decisions thus demand a more in-depth understanding of the contingent relationships among features of the intervention strategies, characteristics of the population, and contextual factors that shape the environment. Furthermore, interdependence among multiple causal effects and the strong likelihood of interference, such that one person's exposure determines many others' outcomes through spillover effects, mean that traditional oversimplifying assumptions are inadequate to study such complex phenomena and predict their outcome distributions with confidence (Marshall & Galea, 2015).

The current study proposes a general model that integrates policy-relevant variables in a unified framework, together with various and distinctive features of the environment and population. By applying this model, policymakers can make better informed decisions and develop effective solutions that reflect the unique characteristics of the population and features of the environment. In the proposed probabilistic contagion model, infection transmission throughout the population is determined by various policy-relevant variables, which operate in a network representing patterns of physical contacts between individuals in the population, as a network structure. The network structure is characterized by two primary, exogenous factors: centrality in its topology and sparsity in its connections. Moreover, to capture heterogeneity at the individual-level, the current research includes a range of individual attributes, such as social distancing behavior, health status, and patterns of physical interactions. Synthesizing the components of this model produces a broader agent-based model (ABM) that can account for heterogeneous individual-level attributes that collectively determine macro-level outcomes.

The stochastic ABM (Bonabeau, 2002; Utomo, Onggo, & Eldridge, 2018) that informs the proposed general framework also can account for major sources of uncertainty and capture heterogeneity in individual-level attributes. In detail, an ABMs seeks to break a system down to its individual components, such as agents, their actions, and their interactions. Simulations of these components in action reveal their effects on the system as a whole (Zheng et al., 2013). Such computational representations of complex systems collect discrete microentities that interact and change over discrete time steps and ultimately give rise to macrosystems Auchincloss & Diez Roux (2008); Miller, Page, & Page (2009). In addition, ABMs produce a rich set of multidimensional data related to macro phenomena, reflecting micro-level agent actions and their dynamic interactions at various levels of temporal and spatial resolutions (Lee et al., 2015). In summary, ABM can depict a complex system through controlled computational experiments. This capability makes them well-suited for the current research context, which features extreme complexity, uncertainty, and causal inference questions, due to a high degree of the interdependence of the underlying mechanisms. Furthermore, logistical infeasibility and ethical concerns prevent randomized controlled trials. Yet advancements in computational technology facilitate the application of ABM, prompting its growing adoption among recent epidemiology studies Aleman, Wibisono, & Schwartz (2011); Epstein (2009); Helbing et al. (2015); Ip et al. (2013); Macal (2016); Marshall & Galea (2015); Tracy, Cerdá, & Keyes (2018).

Ultimately, the objective of the framework is not to offer accurate forecasts or numerical predictions about specific outcomes, on the basis of historical data. Rather, it aims to produce qualitative insights concerning the overall tendency of a complex system in various conditions. Accordingly, the current article offers several stylized examples and reports on an extensive, controlled experiment. The model outputs still demand caution, because they offer solely qualitative evidence (Bonabeau, 2002), applicable for deriving insights and deciding on general policy directions. From this perspective, the findings are suggestive and directional, rather than conclusive.

2. Related literature

A well-known class of compartmental models of infectious disease in epidemiology literature (Kermack & McKendrick, 1927) examines the flow of a population along three phases: susceptible, infected, and recovered, i.e., SIR models.[1] They are deterministic and nonlinear, based on differential equations, and they can describe the spread of a virus, compute the total number of people infected, and estimate various epidemiological parameters, such as the basic reproduction number (Hethcote, 2000). Recent extensions attempt to account for intervention policies and make inferences about their consequences, by examining the effectiveness of quarantine measures Atkeson (2020), considering the role of testing and diagnosis Berger, Herkenhoff, & Mongey (2020), or accounting for social distancing by placing an upper limit on the number of interactions Obiols-Homs (2020). To predict the total number of infected or recovered persons in a population, such models impose specific transmission mechanisms across phases.

In social epidemiology and public health contexts, agent-based modeling instead enables population-level inferences from micro-level rules in simulated populations over time and space (El-Sayed, Scarborough, Seemann, & Galea, 2012; Epstein, 2006; Rahmandad & Sterman, 2008; Tracy et al., 2018). Moreover, whereas contagion models based on differential equations assume homogeneity and perfect mixing within compartments (Mahajan, 2010; Riley, 2007), ABM allows for heterogeneity in individual attributes (Aleman et al., 2011; Rahmandad & Sterman, 2008). However, ABMs also suffer some limitations, which must be acknowledged but that do not undermine its use for the current research. First, these models are cognitively demanding, in terms of the model building process, such that it must take place at the appropriate level of description, with just the right amount of detail (Bonabeau, 2002). Second, ABMs are computationally expensive during the execution

[1] Several variants of the basic SIR model have been developed, such as the SEIR model, which includes an *exposed* state between the susceptible and infected states.

stage, because they require Monte Carlo sampling and multiple model runs to achieve statistical robustness. The result is a massive volume of data that can hinder the efficient discovery of relevant patterns (Lee et al., 2015; Rahmandad & Sterman, 2008). This issue also limits the size of the population that can be included in the analysis to few hundreds, rather than the millions of actors that can be included in differential equation-based models (Rahmandad & Sterman, 2008). Third, the calibration of ABMs is exceptionally challenging, without any consensus regarding the most suitable statistical or ad hoc practices Marshall & Galea (2015); Platt (2020). Despite these limitations, ABMs make causal inferences and simulate counterfactual outcomes in the presence of complexity (Marshall & Galea, 2015); even if the considerable computational costs create a trade-off between realism and quantification of the complex, uncertain phenomena in the study (Helgeson, Srikrishnan, Keller, & Tuana, 2020), simulations still are our best cognitive representation of complex reality and our most profound conceptions of what reality is (Ip et al., 2013; Vallverdú, 2014). That is, with ABMs, researchers can generate outputs that *mimic* those of the phenomenon of interest. With this computational laboratory, then one is able to examine counterfactuals by changing some key conditions of the imitating system Tracy et al. (2018). In turn, ABMs are particularly beneficial when the underlying causal mechanism exhibits a high degree of interdependence and when interference is present, such that one person's exposure to diseases affects other people's outcomes Marshall & Galea (2015). Both of these features are focal elements of the phenomenon examined in the current research.

From an epistemological perspective, simulation in general, and ABMs in particular, provide and an alternative scientific method, beyond deduction or induction. Deduction functions to derive theorems from assumptions; induction seeks patterns in data. In contrast, simulation starts with a set of explicit assumptions, similar to deduction, to generate data, rather than direct measurement of the real world, that are suitable for analysis by induction. Consequently, simulation fundamentally differs from deduction and induction in terms of implementation and goals. Through this juxtaposition, simulation offers enhanced understanding of complex systems, achieved through controlled computational experiments (Axelrod, 2006), including insights into how macroscale dynamics emerge from microscale interactions Auchincloss & Diez Roux (2008).

The model proposed herein relates to the literature on network theory and random graphs as well (Abbe, 2017; Easley, Kleinberg et al., 2010; Gilbert, 1959; Hellmann & Staudigl, 2014), such that it uses insights from these research domains to account for structured interaction patterns among individual actors (Kempe, Kleinberg, & Tardos, 2003). The importance of network structure and its impacts is evident in various contexts, including social contagion, epidemiology, and diffusion (Abrahamson & Rosenkopf, 1997; Davis, 1991; El-Sayed et al., 2012; Manshadi, Misra, & Rodilitz, 2020; Muller & Peres, 2019; Rahmandad & Sterman, 2008). For example, disease transmission depends substantially on people's contact networks, which directly influence their patterns of interactions and thus the transmission rate. If a contact network comprises several connected communities, the pace and scale of the spread still might differ depending on the community-size distribution and population density within each community. Network topology can have a significant impact too (Davis, 1991; Manshadi et al., 2020; Muller & Peres, 2019); the presence of a large central component in the network, such as social hubs (Goldenberg, Han, Lehmann, & Hong, 2009; Muller & Peres, 2019) will lead to a different transmission pattern than would exist if the network is completely decentralized across several fairly equally sized communities. These considerations have clear policy implications, especially for developing context-sensitive solutions.

The framework developed in this article uses a random network model to represent human contact networks. Various types of real networks have been constructed empirically, such as the diffusion of innovation or opinion through social contagion (Iyengar, Van den Bulte, & Valente, 2011). However, for diseases, the process is sufficiently complex and unobservable at the person-to-person level that it is appropriate to model it as random (Easley et al., 2010, pages 568-569). The random network model applied for the current research starts by generating a random community-size distribution based on the integer partitioning problem from number theory (Fristedt, 1993), and by introducing a new algorithm. Next, it constructs networks within each community according to the Erdõs-Rényi random graph model (Erdõs & Rényi, 1960; Gilbert, 1959) and connects the communities on the basis of a stochastic block model (Abbe, 2017).

The intra-community connections that emerge from the Erdõs-Rńyi model reveal a binomial degree distribution within the community. However, recent network science literature questions this property, in that several types of empirical networks can follow different patterns in their degree distribution, such as power-law distributions in scale-free networks (Barabási & Albert, 1999; Barabási, Albert, & Jeong, 2000; Barabási & Bonabeau, 2003). In such networks, the fraction of nodes with degree k may depend on a power law $k^{-\alpha}$, which has broad implications for the network structure. Such patterns ostensibly appear in many types of networks, e.g., online social networks, citation networks, networks of organizations, or biological networks (Albert, Jeong, & Barabási, 1999; Barabási et al., 2000; Newman, 2003b), yet various researchers challenge their underlying assumptions and credibility, on statistical and theoretical grounds (Broido & Clauset, 2019; Willinger, Alderson, & Doyle, 2009). To represent networks of human contacts though, the assumptions of the Erdõs-Rńeyi model combined with stochastic block models appear plausible because heterogeneity in the degree distribution is directly invoked by the dispersion in community-size distribution, as described by previous research as well (Jackson, 2019; Karrer & Newman, 2011; Newman, 2003a) and chapter 2 in (Jackson, 2019). Readers interested in further reviews of these models can find insights about network structures Burt (1980), statistical networks Goldenberg, Zheng, Fienberg, Airoldi et al. (2010), and random networks Jackson (2010); Newman (2018) in prior literature.

3. Model construction

This study uses the following notation:

- $\mathcal{I}$ is a finite set of individuals.
- Time is discrete, with unknown finite horizon, i.e., $t = 0, 1, 2, \cdots, T$.
- $S_i(t)$ is a random variable that represents individual i's state at any time t and can take one of the five possible values: susceptible, infected, detected, recovered, and perished. By default, the neutral state of an individual is susceptible, i.e., $S_i(0) =$ susceptible, that is healthy, or uninfected.[2]
- Exogenous parameter δ is the infection transmission probability of a one-time direct contact between two individuals.
- $N_{ij}(t)$ is an endogenous random variable representing the number of direct contacts between two individuals i and j in a time period t.
- $Z_{ij}(t)$ is an endogenous binary random variable that indicates a transmission of infection from individual j to individual i in time period t.[3].

[2] This state is analogous to a naïve or uninformed individual in the context of new product diffusion in marketing.

[3] Thus, $Z_{ij}(t) = 0$ indicates no transmission of infection from individual j to individual i in time period t

• $\alpha_i \in (0,1)$ is the individual-specific degree of social distancing that individual i adopts, with some heterogeneity in the population, captured by a mixture distribution with a mixing factor π.

• W_i is a random variable representing the incubation period of an infectious agent i, equivalent to the time interval that starts when i is infected and ends when that infection is detected. The parameter ω reflects the testing capacity. with a larger value of ω, the expected incubation period W_i should be shorter.

• H_i is a random variable, for which the distribution parameter η represents the hospitalization period of a detected agent i.

• θ_i is the self-recovery chance, without receiving any medical service, of an infected agent, for which distribution parameters θ_0 and θ_1 reflect the population's overall health.

• The exogenous parameters ϕ and κ are, respectively, the ideal contributions of the healthcare system to the recovery chances of a hospitalized agent, and health service capacity. The endogenous parameter $\phi_{eff}(t)$ is the effective contribution of the healthcare system to the recovery chances of a hospitalized agent, which depends on ϕ, κ, and the number of hospitalized agents at time t.

• $\rho_i(t)$ is the endogenous, individual-specific, time-varying recovery chance of an infected agent discharged from the hospital at time t, which depends on θ_i and $\phi_{eff}(t)$.

• γ, p, and q are parameters of physical contact network, such that γ controls network centrality, and p and q are the within- and between-community link densities of a stochastic block model with $K(\gamma)$ communities.

Starting with a population of size $|\mathcal{I}|$ and $\mathcal{I}_0 \subset \mathcal{I}$ initially infected agents, the central prediction is that every time a susceptible individual i meets an infectious agent j, the infection transmits with a probability δ. For direct contacts, infection transmission occurs according to a Bernoulli experiment with a parameter δ. If two individuals meet multiple times within a specific period, each transmission possibility takes place independently. Formally, if individual i, $S_i(t) = \text{susceptible}$, meets individual j, $S_j(t) = \text{infected}$, for N_{ijt} times in time period t, conditional on $N_{ij}(t)$ and independence, the binary random variable $Z_{ij}(t)$ follows a Bernoulli distribution with parameter $1-(1-\delta)^{N_{ij}(t)}$:

$$Z_{ij}(t)|N_{ij}(t) \sim Bernoulli\Big(1-(1-\delta)^{N_{ij}(t)}\Big). \tag{1}$$

Because the number of direct contacts between two individuals i and j in a time period t, $N_{ij}(t)$, is a random variable, its realized value for each trial depends on whether the two individuals are directly linked in the network of physical contacts, e.g., neighbors or colleagues, and on their patterns of social distancing behavior. Such factors determine the *distance* between two individuals; that is, distance depends on the network structure and social distancing behavior of the two individuals. Distance in this sense represents a latent factor that links social network structure and social distancing behavior to the frequency of direct contacts between two individuals in a unit of time. The construction process of these two components will be discussed in more detail in the next section. At this stage, suppose that the vector $\mathbf{N}_i(t)$, with its jth component representing $N_{ij}(t)$, follows an arbitrary discrete joint probability with mass functions $G_i(t)$. Thus, $G_i(t)$ describes the interaction patterns of i in the network over time. In other words, G provides the main source of differentiation among members of the population when updating the individual's status. The differentiation, as noted, is based on the network structure and social distancing such that these two factors collectively determine the likelihood of contact between an individual i and other members of the population.

For a given t, the $N_{ij}(t)$ Bernoulli trials are independent, and $Z_{ij}(t)$ random variables are independent across j. In other words, the outcome of $Z_{ij}(t)$ is not informative with regard to $Z_{ik}(t)$ for $k \neq j$. Then, the infection risk for i can be obtained as follows:

$$\begin{aligned}
&\mathbb{P}\Big\{S_i(t+1) = \text{infected}\Big|S_i(t) = \text{susceptible}\Big\}\\
&= 1 - \sum_{\mathbf{N}_i(t)}\Bigg[\prod_{j\in\mathcal{I}\setminus i}\Big(\mathbb{I}[S_j(t) = \text{susceptible}]\\
&\quad + (1-\delta)^{N_{ij}(t)}\mathbb{I}[S_j(t) = \text{infected}]\Big)\Bigg]G_i\Big(\mathbf{N}_i(t)\Big),
\end{aligned} \tag{2}$$

where $\mathbb{I}[\cdot]$ is an indicator function that equals 1 if its input argument is true, and 0 otherwise.[4]

Proof: By conditioning on $\mathbf{N}_i(t)$,

$$\begin{aligned}
&\mathbb{P}\Big\{S_i(t+1) = \text{susceptible}\Big|S_i(t) = \text{susceptible}, \mathbf{N}_i(t)\Big\}\\
&= \prod_{j\in\mathcal{I}\setminus i}\Big(1-\big[1-(1-\delta)^{N_{ij}(t)}\big]\mathbb{I}[S_j(t) = \text{infected}]\Big)\\
&\Rightarrow \mathbb{P}\Big\{S_i(t+1) = \text{infected}\Big|S_i(t) = \text{susceptible}\Big\}\\
&= 1 - \sum_{\mathbf{N}_i(t)}\Bigg[\prod_{j\in\mathcal{I}\setminus i}\Big(1-\big[1-(1-\delta)^{N_{ij}(t)}\big]\\
&\quad \times \mathbb{I}[S_j(t) = \text{infected}]\Big)\Bigg]G_i\Big(\mathbf{N}_i(t)\Big).
\end{aligned}$$

By expanding the argument under the product, and considering that $1-\mathbb{I}[S_j(t) = \text{infected}] = \mathbb{I}[S_j(t) = \text{susceptible}]$, Eq. (2) is derived. QED

An infected individual i remains asymptomatic for a random period of T_i before being detected.[5] Once detected though, this individual is separated from the population, e.g., hospitalized or quarantined, for a random period of H_i before recovering or perishing. Recovered individuals remain immune to the infection during the time window of the analysis.

3.1. Social distancing

The frequency of direct contact between two individuals depends on whether they are directly linked in the social network and their social distancing behavior. These two factors combine to constitute the notion of distance, denoted by $d \in \mathbb{R}^+$. Distance, as a medium, inversely relates to the number of direct contacts between two individuals i and j in a single period of time, denoted by λ_{ij}. Assuming independence of the number of direct contacts in non-overlapping periods, by the law of rare events, the number of direct contacts between i and j follows a Poisson distribution,

$$N_{ij}(t)\Big|i \text{ and } j \text{ directly linked } \sim Poisson(\lambda_{ij}), \tag{3}$$

where $\lambda_{ij} \propto 1/d_{ij}^2$.

If the two individuals are not directly linked in the network, then $d_{ij} \to \infty$. For two individuals directly linked in the network, the distance is mainly determined by the degree of social distancing that they adopt, specifically by the largest of these two degrees. This individual-specific parameter is denoted by $\alpha_i \in (0,1)$, where a larger α_i implies a greater degree of social distancing by $i \in \mathcal{I}$. For $i, j \in \mathcal{I}$ that are directly linked in the network, if $\alpha_i \to 0$ and $\alpha_j \to 0$, then $d_{ij} \to 0$. On the other hand, if $max(\alpha_i, \alpha_j) \to 1$, it means that at least one of the two is exercising a strict self-isolation, hence $d_{ij} \to \infty$. Moreover, for identification purposes, the middle point on the scale of α is presumed to correspond, in expectation, to one direct contact in a unit of time, i.e., $\mathbb{E}[N_{ij}(t)] = 1$

[4] Note that the product in Eq. (2) is taken over all the population, because if $N_{ij}(t) = 0$ for some j, then the resulting value would be equal to $\mathbb{I}[S_j(t) = \text{susceptible}] + \mathbb{I}[S_j(t) = \text{infected}] = 1$, which is a neutral element of multiplication.

[5] The stochastic incubation period varies across individuals.

when $max(\alpha_i, \alpha_j) = 1/2$. Following this standardization, the scaling factor between the Poisson parameter and distance vanishes, and $\lambda_{ij} = 1/d_{ij}^2$. The following specification for d satisfies these properties:[6]

$$d_{ij} = \frac{\max(\alpha_i, \alpha_j)}{1 - \max(\alpha_i, \alpha_j)}. \tag{4}$$

Individuals in the network do not exhibit the same degree of social distancing behavior.[7] [8] To capture this heterogeneity, parameter α is modeled as a finite mixture of Beta distributions with support $(0, 1)$. The parameters of the components of this mixture distribution are defined in such a way that each corresponding density function is either strictly increasing or decreasing. Then by varying the model parameters, it is possible to shift individual degrees of social distancing towards one of the extremes of the support range without a mass concentration in the vicinity of the middle point, which provides a direct control for heterogeneity in social distancing behavior. Specifically,

$$\alpha_i \sim \pi \text{Beta}(a_c, 1) + (1 - \pi)\text{Beta}(1, a_{nc}), \tag{5}$$

where $0 \leq \pi \leq 1$ is the mixing factor and represents a fraction of the population that exercises a great degree of social distancing, e.g., conforming group, including a within-group variation controlled by $a_c \in \mathbb{Z},\ a_c > 1$. The second component, with a population fraction of $1 - \pi$, represents the sub-population that does not adopt social distancing behavior, such that $a_{nc} \in \mathbb{Z},\ a_{nc} > 1$ is the shape parameter of the distribution. The density function of the first (second) component is strictly increasing (decreasing) in a_c (a_{nc}). By varying a_c and a_{nc}, it is possible to model social distancing behavior within each group, and then π controls for the social conformity level with social distancing behavior, such that a greater value of π implies more collaboration among the population in terms of exercising social distancing.[9]

3.2. Stochastic incubation and hospitalization

An essential characteristic in the spread of many types of infectious diseases, including COVID-19, is the presence of a significant incubation period.[10,11] The incubation period differs among people, depending on various factors such as their demographics and health status. Expanding testing capacity can shorten this period, though, because more extensive testing helps detect and isolate asymptomatic infectious agents, which curbs the virus's spread.

The incubation period W_i of the infectious agent i is the time difference between infection and detection. An infectious agent is detected according to a Bernoulli trial at the end of each time period. Therefore, W_i follows a geometric distribution,

$$W_i | i \text{ infected } \sim Geometric(\omega), \tag{6}$$

where the parameter ω reflects the testing capacity.[12] In general, raising the testing capacity by a factor of n changes the distribution parameter in Eq. (6) from ω to $1 - (1 - \omega)^n$.[13]

Similarly, an infected individual i, when detected, is hospitalized for a period H_i. Similar to the incubation period, H_i differs across the population,

$$H_i | i \text{ detected } \sim Geometric(\eta). \tag{7}$$

3.3. Recovery chance and healthcare system overload

At the end of the hospitalization period, or detection time plus H_i, an agent i is recovered or perished.[14] The individual-specific, time-varying, recovery chance of a detected person discharged from the hospital at time t, $\rho_i(t)$, then can be decomposed across its two dimensions, individuals and time, by the individual-specific general health status of the person, denoted by θ_i, and the time-dependent effective contribution of the healthcare system that provides capacity-constrained medical care, denoted by $\phi_{eff}(t)$. The first component, θ_i, indicates the recovery chance of the individual without receiving any health service, so if $\theta_i \sim \text{Beta}(\theta_0, \theta_1)$, the parameters θ_0 and θ_1 reflect the overall health status of the population.

The second component augments these self-recovery probabilities by $\phi \times (1 - \theta_i)$, assuming an ideal situation in which the healthcare system is not overloaded. The parameter ϕ represents such ideal healthcare service quality and can be understood as the percentage reduction in mortality risk of an infected agent who

[6] This specification relies solely on the largest of the two social distancing parameters when determining the distance between two individuals, so for example, it cannot differentiate between $(\alpha_i, \alpha_j) = (0.1, 0.9)$ and $(0.8, 0.9)$. Nevertheless, distance in the latter case might be slightly smaller. For this reason, in the simulation computations presented herein, the *max* function is replaced by the smooth approximation $max(\alpha_i, \alpha_j) = (\alpha_i e^{n\alpha_i} + \alpha_j e^{n\alpha_j})/(e^{n\alpha_i} + e^{n\alpha_j})$, where n is a smoothing factor, and the approximation error becomes smaller as n grows. In the computational analysis, $n = 10$ which provides an accurate approximation of the *max* function within the range of parameters values in the experiment.

[7] Such heterogeneity can have a large impact on the overall outcome of the system and has direct implications for policymaking. The effectiveness of a response to the virus spread might largely rely on effective collaboration among individuals. Therefore, it is important to examine how the outcomes vary by the level of heterogeneity in individuals social distancing behavior.

[8] A good proxy for such heterogeneity may be income differences across the population (Chiou & Tucker, 2020), measured with macro-level indicators such as the Gini index (Van den Bulte & Stremersch, 2004). Further empirical investigations are needed to establish a more comprehensive characterization of this component.

[9] The density function for α_i, given π, is $f_{\alpha_i|\pi}(x) = \pi a_c x^{a_c - 1} + (1 - \pi) a_{nc}(1 - x)^{a_{nc} - 1}$. By deriving the first and second moments of α_i, it can be shown that $\mathbb{E}[\alpha_i] = \pi \mu_c + (1 - \pi)\mu_{nc}$, and $Var(\alpha_i) = \pi \sigma_c^2 + (1 - \pi)\sigma_{nc}^2 + \pi(1 - \pi)(\mu_c - \mu_{nc})^2$, where $\mu_c = \frac{a_c}{a_c + 1}$ and $\sigma_c^2 = \frac{\mu_c(1 - \mu_c)}{a_c + 2}$ are, respectively, the mean and variance of the first component of the mixture distribution. Similarly $\mu_{nc} = \frac{1}{a_{nc} + 1}$ and $\sigma_{nc}^2 = \frac{\mu_{nc}(1 - \mu_{nc})}{a_{nc} + 2}$ are the mean and variance of the second component, respectively.

[10] This is related to the concept of time-delay in diffusion and social contagion, which can introduce a memory element to the contagion process and produce non-Markovian characteristic.

[11] In the case of COVID-19, many people become symptomatic only after several days of being transmitters, and some others remain asymptomatic for their full infection period.

[12] The choice of a geometric distribution is motivated by the analogous convention of using Poisson distribution to represent the number of events in a time interval in the discrete-time events and complex systems simulations, which is based on the postulate that the number of events occurring in non-overlapping time intervals are independent. This is a suitable property, especially in the context of this research, since assuming otherwise requires a priori information to impose a particular dependence structure between the occurrence of the events. Under the assumption of Poisson distribution, the time interval between two consecutive events follows an exponential distribution which posits the interesting property of memorylessness. It turns out the exponential distribution is the only continuous distribution exhibiting this property. In other words, if a continuous random variable has the memorylessness property, then it necessarily follows the exponential distribution. The interpretation of this property in the context of the spread of infectious disease is the same as that of Poisson distribution. Specifically, knowing that an agent has not been detected till today does not affect, ceteris paribus, the probability of her/him being detected in the next day. Similar to the independence property of Poisson distribution, memorylessness of the exponential distribution is a suitable property in the context of this paper, as well as simulation of other complex systems, because assuming otherwise requires explicit knowledge of the underlying mechanism, which usually is not available. The exponential distribution is defined for continuous random variables, whereas time is treated as discrete in this paper. Being the only discrete distribution that satisfies the memorylessness property, the geometric distribution is the discrete analogous of the exponential distribution.

[13] One way to interpret the parameter ω is to imagine that there is a test center that, in each period, randomly selects some individuals for the test, each with a probability ω. With the assumption that there are no false positive or false negative errors in the test results, doubling the testing capacity yields two identical centers that operate independently and select individuals with replacement, which then would change the parameter of the geometric distribution to $1 - (1 - \omega)^2$.

[14] Recovered people are no longer susceptible to the virus during the analysis time window, so it could lead to herd immunity as occurs when the immune proportion exceeds $(1 - 1/\mathcal{R}_0)$, where $\mathcal{R}_0$ is the reproduction number and represents the average number of people to which a single infectious agent transmits the virus. An epidemic with a reproduction number below 1 gradually disappears.

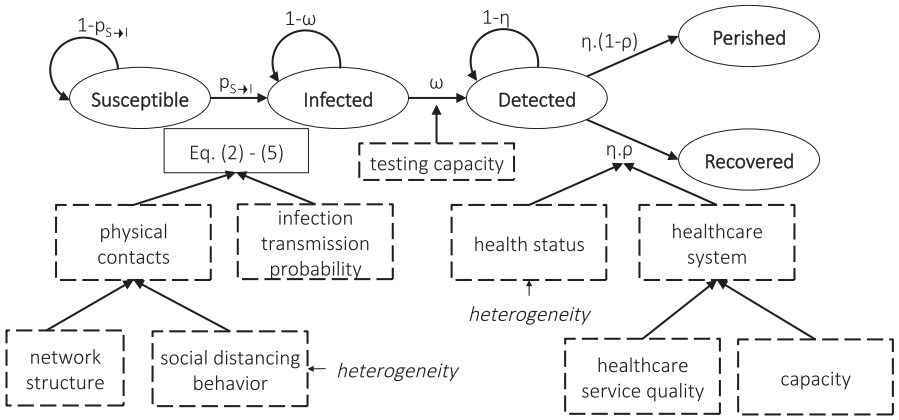

Fig. 1. Summary of the model in form of a state transition diagram and the main factors affecting the transition probabilities. The ellipses represent the states; dashed boxes are the influential factors affecting transition probabilities; $P_{S\to I}$ is the transition probability from susceptible to infected, as described in Equations (2)-(5).

receives health service in the absence of a system overload.[15] Because the healthcare system is capacity-constrained, this health service contribution diminishes as a result of the system overload. Specifically, assuming a capacity of κ, in terms of hospital beds per inhabitant, ϕ is linearly discounted with respect to demand overload:

$$\mathbb{P}\Big\{S_i(t) = \text{recovered } |S_i^{-1}(\text{detected}) = t - H_i\Big\} = Bernoulli\Big(\rho_i(t)\Big), \tag{8}$$

in which

$$\begin{aligned} \rho_i(t) &= \theta_i + \phi_{eff}(t)(1-\theta_i), \\ \phi_{eff}(t) &= \frac{\phi}{\max\Big\{\frac{\sum_{i\in\mathcal{I}} \mathbb{I}[S_i(t) = \text{ detected }]}{\kappa\cdot|\mathcal{I}|}, 1\Big\}}. \end{aligned} \tag{9}$$

Due to its adaptive property, health service quality thus contributes to the recovery rate through two main factors: service quality and service capacity, which the above formulation can disentangle. As a strategic decision, policymakers might decide to improve health service quality (ϕ), such as by investing in new medical equipment and medical training programs, or increase the healthcare capacity (κ), such as by constructing additional hospitals to cope with demand shock without compromising health service quality.[16,17]

The state transition diagram in Fig. 1 depicts the proposed model by highlighting the major factors influencing the state transition probabilities are highlighted. The process is not time-homogeneous because the state transition probabilities vary as time unfolds and the system evolves. Holding all other influential factors constant, the probability of transition from susceptible to an infected state increases for a person when more other members of the contact network are infected, whereas it decreases if a greater proportion portion of the population enters recovered or perished states. Similarly, an agent's recovery chance decreases as more other agents enter the detected state.

4. Network topology

Physical contact networks can be modeled as random networks, consisting of communities of different sizes, for which intra-community densities are greater than inter-community densities. Constructing such networks involves three main steps. First, the number of communities, denoted by K, and a community-size distribution, where i_k denotes the kth community size, are randomly generated according to a proposed random integer partitioning algorithm. Second, each community, given its randomly generated size, is constructed according to the Erdős-Rényi random graph model. Third, for the construction of the overall network, the communities are linked according to a stochastic block model, specifically, the assortative planted partition model. The resulting network of contacts can be characterized by variation in the community-size distribution, or the dispersion of i_k values, and the within-community link density. To avoid generating communities with absurdly small sizes, a lower bound is imposed on the i_k values. To generate a random partition of $|\mathcal{I}|$ that satisfies these properties, a probabilistic algorithm is used.

Specifically, the objective of the first step is to divide $|\mathcal{I}|$ into some positive integers, $i_1 \geq i_2 \geq \cdots \geq i_K$, where the number of parts representing the total number of communities K is itself a random variable. By definition $|\mathcal{I}| = \sum_{k=1}^{K} i_k$. For this classical NP-hard combinatorial optimization problem, I propose a random integer partitioning algorithm.

Suppose there are $|\mathcal{I}|$ baskets, labeled 1 to $|\mathcal{I}|$, each containing a single ball. Given the parameter values for the variation in the community-size distribution and lower bound on i_k, the aim is to redistribute the balls among the baskets so that the number of non-empty baskets represents the number of communities K, and the number of balls in each basket represents the size of the corresponding community, i_k. This is performed based on the following stages. In the first stage, each ball in the baskets 2 to $|\mathcal{I}|$ is an independent candidate for transfer to basket 1, with probability γ. Following such potential transfers, basket 1 is closed. [18] Then in the second stage, each remaining ball in baskets 3 to $|\mathcal{I}|$ is an independent candidate for transfer to basket 2, with probability γ, and basket 2 is closed at the end of this process. The process continues until all remaining open baskets are empty. The empty baskets

[15] In other words, $\phi = \frac{(1-\theta)-(1-\rho)}{1-\theta}$ in the absence of a system overload. If an infected person's mortality risk is 30%, $\theta = 0.70$, then receiving health service of a quality $\phi = 0.8$ reduces this mortality risk by 80%, or from 30% to $30 - 0.8 \times 30 = 6\%$, that is $\rho = 0.94$.

[16] Note that $\kappa \cdot |\mathcal{I}|$ is the demand that can be handled without compromising health service quality.

[17] Healthcare capacity κ is not necessarily equivalent to the total physical capacities of hospitals because, in reality, not all detected cases are hospitalized. Therefore, the parameter is construed as an adjusted value, relative to the fraction of the detected cases that require hospitalization. If in a city, for instance, only 5% of the detected cases are hospitalized, then the value of the parameter κ reflects the total physical capacity, multiplied by 20.

[18] Therefore, basket 1 contains, in expectation, $1 + \gamma(|\mathcal{I}| - 1)$ balls at the end of this stage.

are removed and the closed ones are sorted, in a descending order, according to the number of balls they contain. The sorted baskets are labeled [1], [2], $\cdots$, $[n]$, $n \leq |\mathcal{I}|$.

Now let $i_{\min}$ be the a lower bound on the admissible parts, such that constraint $i_k \geq i_{\min}$, $\forall k$. In other words, $i_{\min}$ represents the smallest admissible size for a randomly generated community. Then, if according to a comparison of the number of balls in basket $[n]$ with $i_{\min}$, the constraint is satisfied, such that $i_{[n]} \geq i_{\min}$, the process terminates. Otherwise, all the balls in basket $[n]$ are transferred to basket $[n-1]$, basket $[n]$ is dismissed, and the remaining baskets are sorted and labelled again. The process repeats for the basket labelled $[n-1]$ now. The process continues until all baskets contain at least $i_{\min}$ balls. At this stage, the algorithm terminates.

The procedure is described in Algorithm 1. The input for this algorithm thus includes the population size, network centrality parameter γ, and minimum community size $i_{\min}$, and it provides the number of communities and community-size distribution, in accordance with the constraints on the community minimum size and $\sum_{k=1}^{K} i_k = |\mathcal{I}|$.[19]

Finally, to construct the entire network, a random adjacency matrix A of size $|\mathcal{I}| \times |\mathcal{I}|$ is generated, where $A[ij] = 1$, $i < j$, indicates that individuals i and j are directly linked in the network.[20] The matrix A is divided into K blocks where its first i_1 rows and columns correspond to the individuals in the first community, the next i_2 rows and columns correspond to the individuals in the second community, and so on. The submatrix A_k, corresponding to community k with size i_k, has an Erdõs-Rény distribution, with $A_k[ij] \overset{\sim}{iid} Bernoulli(p)$, $i < j$. In other words, any two individuals within the community are linked according to an independent Bernoulli trial with a parameter p. Two individuals i and j from two different communities instead are directly linked according to a Bernoulli trial with parameter q, where $q < p$. Fig. 2 illustrates the generated networks with centrality parameters $\gamma = 0.7$ and 0.3, by setting $p = .20$ and $q = p/100$ or $q = p/20$ for the same generated community-size distribution.

Algorithm 1 Random integer partitioning with restricted dispersion.

procedure CommunityGeneration
Input: $|\mathcal{I}|$, γ, $i_{\min}$
Output: K, $i_1 \geq i_2 \geq \cdots \geq i_K \geq i_{\min}$
  $X \leftarrow \text{ones}(|\mathcal{I}|)$
  for all $j < i \leq |\mathcal{I}|$ **do**
    $r \leftarrow Rand(0,1)$
    $X[i] \leftarrow X[i] + \mathbb{I}[r < \gamma] * (X[j] > 0)$
    $X[j] \leftarrow X[j] - \mathbb{I}[r < \gamma] * (X[j] > 0)$
  end for
  $X \leftarrow sort(X,\ descend)$
  $X[X == 0] \leftarrow [\ ]$
  while $X[end] < i_{\min}$ **do**
    $X[end-1] \leftarrow X[end-1] + X[end]$
    $X[end] \leftarrow [\ ]$
    $X \leftarrow sort(X,\ descend)$
  end while
  $X \leftarrow sort(X,\ descend)$
  $K \leftarrow |X|$, $(i_1, \cdots, i_C) \leftarrow X$
  Return $K, i_1, \cdots, i_K$
end procedure

5. Experimental analysis

Section 5.1 presents a prototype analysis, in the form of a series of sequentially related examples, which provide three relevant demonstrations of the model's (1) flexibility for capturing and disentangling the effects of various components of the complex system, (2) expected outcomes, and (3) ability to support what-if analyses to inform policy analysis and decision making. Then in Section 5.2, an extensive computational experiment is presented which examines and tests for interrelationships among the characterizing parameters of the complex system and its outputs, which provide qualitative insights that can inform policy decisions across a range of settings.

5.1. Prototype analysis

Consider a population of $|\mathcal{I}| = 1000$ individuals, distributed in a decentralized network $\gamma = 0.3$, with $i_{\min} = |\mathcal{I}|/100$, densely connected within-community individuals $p = .50$, and sparsely connected communities $q = p/100$. In terms of its behavior, the population exercises a great degree of social distancing and a relatively low level of within-group variation, so a large value of the parameter a_c in Eq. (5). Firs, let us assume that a high degree of social conformity exists, $\pi = 1$. Subsequent assessments in the same generated network allow for a comparison of the results in the presence of some heterogeneity, such that $\pi < 1$, depicting a scenario where a small fraction of the population breaches ($\pi = 0.95$, $a_{nc} = 5$). At this stage, assume $\alpha_i \sim \text{Beta}(5,1)$, such that on average, there are $\lambda \approx 0.03$ physical contacts in a single time period between two directly linked individuals in the network, and ≈ 0.28 number of interactions on average for an individual in the 95th or greater percentile.[21] In terms of population characteristics, suppose that $\theta \sim \text{Beta}(8,2)$, representing an 80% chance of self-recovery[22], in expectation, with moderate heterogeneity, and $\eta = 0.5$, so any detected agent likely is hospitalized for two weeks. In terms of contagion characteristics, assume that $\delta = 0.2$, so there is a 20% risk of transmission through direct contact. Finally, assume that the healthcare system offers a high level of healthcare quality $\phi = 0.8$, a large healthcare capacity $\kappa = 50$, and a relatively low testing capacity $\omega = 0.30$, so it would take more than 3 time periods, in expectation, to detect an infectious agent. For the initial condition, at time $t = 0$ there are 5 infectious agents in the population. The computations continue until no infected or hospitalized cases remain. As a result, the epidemic duration T, is determined endogenously.

By holding the generated network constant and repeating the simulation several times, it is possible to compute various types of outputs while also quantifying their uncertainty levels to account for inherent stochasticity in the system. Through 50 replica-

[19] The parameter γ controls for variation in the community-size distribution. For $|\mathcal{I}| = 1000$ and $i_{\min} = 10$, for example, two generated instances of $(i_1, \cdots, i_K)$ by Algorithm 1 when $\gamma = 0.7$ are (725, 191, 56, 17, 11) and (720, 193, 54, 33), whereas if $\gamma = 0.3$, the instance would be (318, 176, 140, 98, 72, 60, 43, 33, 20, 19, 11, 10).

[20] A link between two individuals does not necessarily imply friendship or any specific association, other than the possibility of physical contact. Two individuals can be linked in the network because they visit the same supermarket in a neighborhood or use the same bus, two regular situations that can enable direct virus transmission.

[21] Precisely, define $F_\lambda(x) = \mathbb{P}(\lambda \leq x)$ and $Q_\lambda(p) = \inf\{x : F_\lambda(x) \geq p\}$. Then, $\mathbb{E}\ (\lambda \mid \lambda \geq Q_\lambda(0.95)) \approx 0.28$. The number of physical contacts between two individuals i and j in a time period t, $N_{ij}(t)$, is determined by the latent distance d_{ij} that in turn depends on $max(\alpha_i, \alpha_j)$. Assuming that the two individuals are directly linked in the network, the expected value of $N_{ij}(t)$ can be approximated numerically using the distribution of $max(\alpha_i, \alpha_j)$ Sculli & Wong (1985). Specifically, suppose that $\alpha = \max(\alpha_i, \alpha_j)$. In that case $1/d_{ij} = (1-\alpha)/\alpha$, and $\mathbb{E}(\lambda_{ij}) = \mathbb{E}(1/d_{ij}^2) = 1 + \mathbb{E}(1/\alpha^2) - 2\mathbb{E}(1/\alpha)$. These expectations, and other quantities with respect to λ_{ij}, can be approximated numerically using Monte Carlo simulation and drawing from the distribution of $max(\alpha_i, \alpha_j)$, which is known under the iid assumption.

[22] The infamous Imperial College model assumed a value of 0.1 for this parameter, which yielded a largely overestimated prediction of deaths in the United States and the United Kingdom– see the New York Times report Landler & Castle (2020).

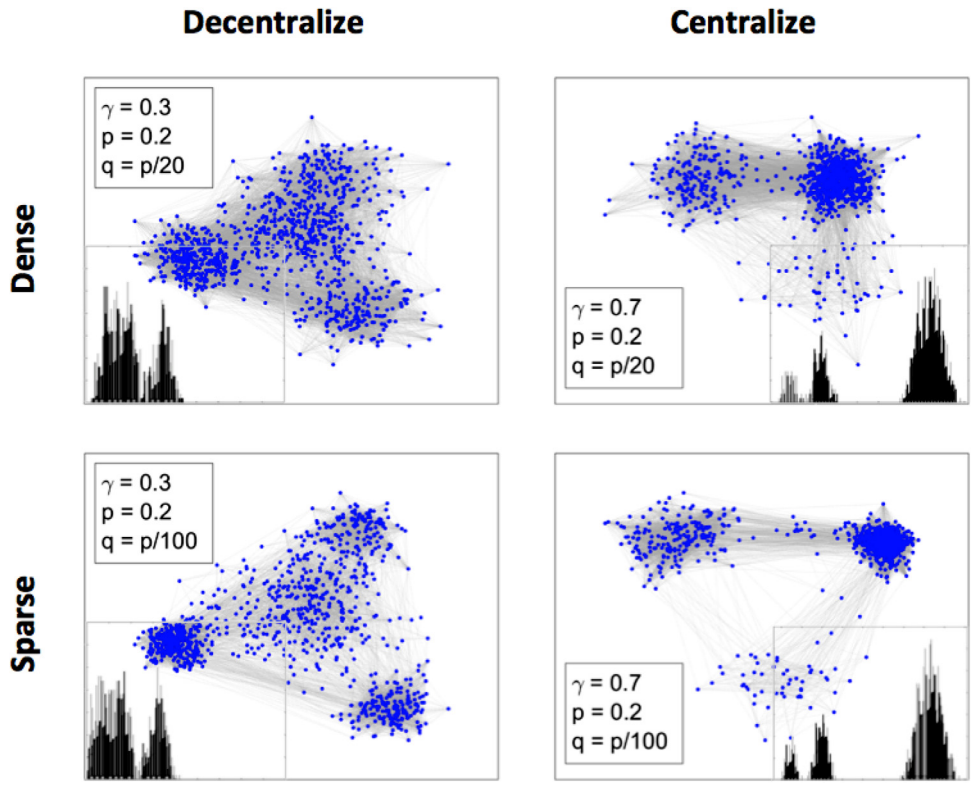

Fig. 2. Constructed random networks using Algorithm 1 for different network structures (left: decentralized – $\gamma = 0.3$, right: centralized – $\gamma = 0.7$), different levels of inter-community connections (top: densely connected communities – $q = p/20$, bottom: sparsely connected communities – $q = p/100$), and fixed intra-community intensity ($p = .20$). The histograms represent the degree distribution of the corresponding random network, which clearly differ from the scale-free networks..

tions of the computations for the above setting, the key outputs, such as the fraction of infected or perished people or epidemic duration, get measured repeatedly in each iteration. In each simulation iteration, the total number of infected, detected, recovered, and perished cases also is stored for each period. Moreover, the number of newly infected cases at each time, as well as the total number of active cases, which are infected but not detected, are stored. The basic reproduction number ($\mathcal{R}_0$), equals to the average number of agents to whom a single infected person transmits the virus, is calculated and stored for each period too. As widely used in epidemiology, this metric helps track the pace of the virus's spread. Finally, the epidemic recovery pace, defined as the time between the epidemic peak and full recovery, with reference to the total epidemic duration, is measured for each setting to indicate the relative point in time corresponding to the maximum load on the healthcare system. To enhance its practical relevance, a peak time in the analysis also is identified, using detected instead of infected cases, which is observable while the other is not. Specifically, the peak time is $t_{peak} = \text{argmax}_t\big[D(t) - D(t-1)\big]$, $t = 1, 2, \cdots, T$, where $D(t)$ is the total number of detected cases in t. Subsequently, the ratio $\nu = t_{peak}/T \times 100$ indicates the recovery speed of the system. Therefore, $\nu = 50$ would imply that the recovery time is equivalent in length to the epidemic duration from its start to its peak, whereas $\nu = 20$ indicates the recovery duration is 4 times longer than the start to the peak of the epidemic duration. Therefore, for a fixed pandemic duration, a greater ν indicates a more delayed peak. For all these measures, nonparametric 99% confidence intervals are computed by bootstrapping, using the bias-corrected and accelerated percentile method and 10,000 bootstrap samples. The results obtained through these analyses are presented in Fig. 3– top row (note the different right-side vertical axis units for perished cases).

In the setting described above, the epidemic is predicted to last, in expectation, $T = 42.8$ periods (with a 99% confidence interval of 38.6 to 46.8) and affect 45.4% of the population (with a 99% confidence interval of 39.1 to 47.7), resulting in the deaths of 2.7% of the population (with a 99% confidence interval of 2.3 to 3.0). The recovery pace is as small as 0.27 (with a 99% confidence interval of 0.24 to 0.30). Even though the basic reproduction number barely exceeds its critical value of 1, the combination of a considerable incubation period and low testing capacity leads to an outbreak.

Next, holding all the characterizing parameters of the setting and the generated random network constant, but assuming some heterogeneity in social distancing behavior, such that 5% of the population breaches ($\pi = 0.95$), produces setting 2. Applying the same analysis, the results indicate a moderate increase in the total number of infected cases ($\sim 15\%$) but a dramatic increase in the number of deaths ($\sim 50\%$).[23] The epidemic duration shortens ($\sim 0.18\%$), but the recovery pace remains unchanged. Increasing testing capacity fourfold, such that $\omega = 1 - (1 - 0.3)^4 = 0.76$, leads to setting 3, which compensates for the small heterogeneity introduced in the number of deaths and brings this number down to 2.5 (99% confidence interval 2.1 to 2.8), close to the full social conformity results in the setting 1. Moreover, this change reduces the number of infected cases (30% versus 45.4% in setting 1), shrinks the epidemic period (19.2 versus 42.8 in setting 1), and slightly increases the recovery pace. In the detailed descriptions in Table 1 the parameter changes at each stage are highlighted. Then Table 2 provides the results obtained from each setting.

[23] To attribute the changes in the outcomes to the change in the parameter value, before simulating the system, I restore the random number generator seed, so the two scenarios and all the generated random numbers are identical in all aspects except for the modified parameter value.

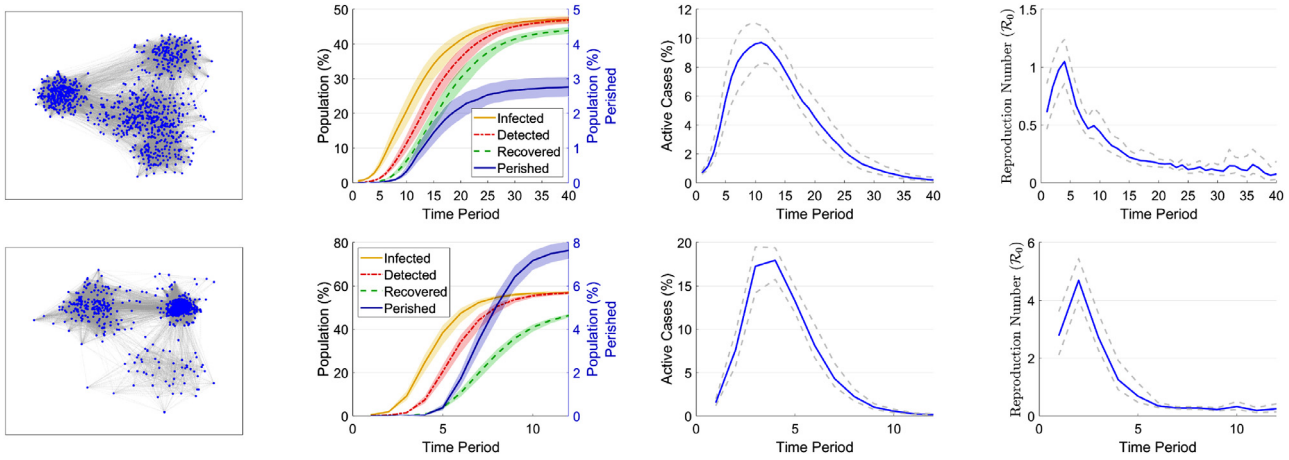

Fig. 3. Results of the simulation corresponding to setting 1 (top) and setting 4 (bottom) in the prototype analysis. The leftmost figure represents the generated random network, held constant in the 50 replications. The other figures depict the state of the network over time and represent, respectively from left to right, the population fraction that is infected, detected, recovered, or perished in each period (note the different right-side vertical axis units for the perished cases); the population fraction that is active, as undetected infectious agents, in each time; and the basic reproduction number at each time. Shaded areas and grey dashed line intervals represent 99% bootstrap confidence intervals, computed using the bias-corrected and accelerated percentile method, with 10,000 bootstrap samples.

Table 1
Parameter values of the sequentially related settings, identical except for the highlighted (bold) values, in the prototype analysis ($|\mathcal{I}| = 1000$).

	Network structure			Infection	Health status			Healthcare system			Social distancing		
Setting	γ	p	q	δ	θ_0	θ_1	η	ω	ϕ	κ	a_c	a_{nc}	π
1	0.3	0.50	$p/100$	0.2	8	2	0.5	0.30	0.8	50	5	5	1
2	0.3	0.50	$p/100$	0.2	8	2	0.5	0.30	0.8	50	5	5	**0.95**
3	0.3	0.50	$p/100$	0.2	8	2	0.5	**0.76**	0.8	50	5	5	0.95
4	**0.7**	0.50	$p/100$	0.2	8	2	0.5	0.76	0.8	50	5	5	0.95
5	0.7	0.50	$p/100$	0.2	8	2	0.5	0.76	0.8	**100**	5	5	0.95
6	0.7	**0.25**	$p/100$	0.2	8	2	0.5	0.76	0.8	100	5	5	0.95
7a	0.7	0.25	$p/100$	0.2	8	2	0.5	0.76	0.8	100	5	5	**1**
7b	0.7	0.25	$p/100$	0.2	8	2	0.5	0.76	0.8	100	5	5	**0.80**

Table 2
Results of sequentially related settings, identical except for the parameter in the condition change column, in prototype analysis ($|\mathcal{I}| = 1000$). The 99% confidence intervals are computed by bootstrapping and using the bias-corrected and accelerated percentile method, with 10,000 bootstrap samples.

		Infected (%)			Perished (%)			Epidemic duration T			Recovery pace ν		
Setting	Condition change	Mean	99% Conf. Int.		Mean	99% Conf. Int.		Mean	99% Conf. Int.		Mean	99% Conf. Int.	
1	–	45.4	39.1	47.7	2.7	2.3	3.0	42.8	38.6	46.8	27	24	30
2	π : 1.00 → 0.95	51.8	50.8	52.7	4.0	3.7	4.3	36.3	34.6	38.5	27	25	30
3	κ : 0.30 → 0.76	30.0	26.1	31.4	2.5	2.1	2.8	19.2	18.0	20.5	34	30	39
4	γ : 0.3 → 0.7	57.1	56.3	57.8	7.7	7.4	8.2	18.0	17.1	18.9	27	25	29
5	κ : 50 → 100	58.2	57.5	59.2	5.9	5.6	6.2	17.4	16.8	18.3	26	25	28
6	p : 0.50 → 0.25	43.2	42.0	44.2	3.0	2.8	3.3	18.0	17.2	18.8	29	27	31
7a	π : 0.95 → 1.00	37.0	32.7	38.2	2.1	1.8	2.3	19.0	17.7	20.1	34	31	36
7b	π : 0.95 → 0.80	55.6	54.9	56.4	5.6	5.3	5.9	16.4	15.5	17.5	24	22	26

In the next stage, to project setting 3 to a centralized network, all the parameter values remain constant except that γ changes from 0.3 to 0.7, producing setting 4. In a centralized network, this configuration produces a very different picture than its decentralized counterpart. Specifically, changing the parameter value reveals a dramatic increase in the number of infected and perished cases, which doubled and triple, respectively. Nevertheless, the epidemic duration remains unchanged. The comparisons depicted in bottom row of Fig. 3 indicate that the number of active, undetected, infectious agents grows faster and reaches a higher peak in the centralized network, as is also reflected by the values of the basic reproduction number over time. These results demonstrate the prominent role of network structure. The variations in outputs due to stochasticity in agents' behavior are marginal (narrow confidence intervals), but variations due to changes in the network structure are of an order of magnitude.

In settings 5 and 6, the goal is to explore different intervention alternatives, one for each stage, that might mitigate the adverse impacts introduced due to the increased network centrality. In the first stage, healthcare capacity κ, is doubled from 50 to 100. Doing so lowers the number of deaths by about 23% but has a negligible impact on the number of infected cases, epidemic duration, or recovery pace. Greater numbers in a centralized network might be attributed to shorter path lengths between pairs of individuals, due to the presence of a "hub", so weakening connections in the network could be an effective response. Constant decontamination of surfaces and public areas would decrease the rate of mediated virus transmission by eliminating the route through phys-

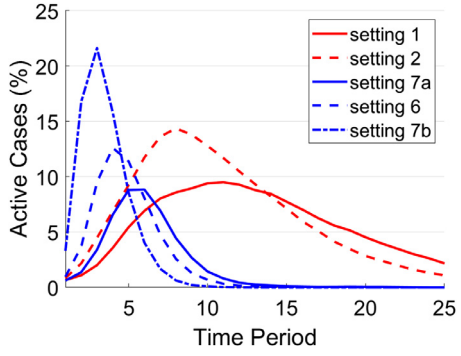

Fig. 4. Impact of surveillance in decentralized (red) and centralized (blue) networks. The graph depicts the mean value of active cases over time for settings 1 and 2, with red solid and dashed lines, respectively, corresponding to $\pi = 1$ and $\pi = 0.95$ in the decentralized network; as well as settings 7*a*, 6, and 7*b*, with solid, dashed, and dash-dotted lines, respectively, corresponding to $\pi = 1$, 0.95 and 0.80 in the centralized network. The red and blue lines should not be compared directly, because they differ in terms of other parameters. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ical contact with common objects and surfaces. This intervention accordingly can be reflected in the parameter related to the intra-community density p. Changing p from 0.50 to 0.25 decreases the number of infected cases, as expected, and also substantially reduces the number of deaths ($\sim 50\%$).

In settings 7*a* and 7*b*, changes to the degree of surveillance are explored. Both settings are produced by changing a parameter in setting 6. First, surveillance is tightened by enforcing maximal conformity to social distancing, updating π from 0.95 to 1 (setting 7*a*). Second, a further de-escalation of social distancing restrictions changes π from 0.95 to 0.80 (setting 7*b*). In contrast with the decentralized environment, tightening surveillance exhibits no impact on the epidemic duration but increases the recovery pace. The implication is that minor heterogeneity in social distancing can shift the peak time to the left in both centralized and decentralized networks, but the shift in the former stems from the diminishing recovery pace. In the decentralized network, it results from shortening the epidemic duration, as depicted in Fig. 4.

These results highlight the interrelated, complex nature of the interactions between the contextual factors and policy parameters, as well as the nonlinear behavior of the system. Yet, for this assessment, the generated random network for each setting was held constant. The narrow confidence intervals demonstrate minor variations in measured outputs that can arise due to the agents' stochastic behavior and attributes. Nevertheless, to enhance the external validity of the results, it is necessary to account for randomness in the generated network, given the network structure parameters. Therefore, in the next section, each setting is repeatedly analyzed in the same way but across a range of the network's randomized manifestations.

5.2. Simulations

This section presents an extensive computational experiment, designed to reveal interrelationships among the characterizing parameters of the system and its outputs and thus produce qualitative insights to inform policy decisions across a range of settings. From this perspective Haber et al. (2007), the objective is to draw suggestive and directional conclusions, rather than numerical predictions. Therefore, a baseline condition is defined for each network structure. The baseline serves as a reference against the outputs of different settings with the same network parameters, such that it reveals the *improvements* in the respective output measures due to the various intervention methods, rather than absolute values per se. These differences in the values are stored for further analysis. Moreover, this instrument supports comparisons of the relative impacts of policy parameters *across* network settings, in that taking the differences cancels out the uncontrolled idiosyncratic effects of a randomly generated network in each instance of a scenario. In turn, this measurement strategy enables the identification of the intervention effect parameters.

The design parameters of the experiment and their corresponding levels are presented in Table 3. Each network structure is characterized by the triple $\langle \gamma, p, q \rangle$ for which the baseline condition is defined as $(\theta_0, \theta_1) = (5, 5)$, $\omega = 0.30$, $\phi = 0.40$, $\kappa = 50$, and $\pi = 0.80$. [24]

Throughout the analysis, the population size is fixed at $|\mathcal{I}| = 1000$. The experiment proceeds through following steps:

Step 1: Specify γ and $i_{\min} = |\mathcal{I}|/100$, and generate the number of communities K, and community-size distribution, $i_1, \cdots, i_K$, according to Algorithm 1. Specify p and q.

Step 2: Generate a random network based on the stochastic block model. Hold the generated network constant in steps 3 and 4.

Step 3: Set π and generate individual-specific social distancing parameter values α_i by drawing from the distribution in Eq. (5). Set θ_0 and θ_1 and generate individual-specific self-recovery chance parameter values θ_i. Set ω and η and generate incubation and hospitalization period values for each individual, W_i and H_i, respectively. Set ϕ, health service quality in the absence of demand overload; and κ, healthcare system capacity. Simulate the system and compute the total fraction of infected and perished cases, as well as epidemic duration and recovery pace.[25] Hold the generated network constant and repeat the analysis 50 times, storing the mean values of the respective outputs over the 50 replications.

Step 4: Repeat step 3 for all the configurations defined by updating the levels of the parameters. Compute and store differences in the measured outputs with reference to the baseline condition.

Step 5: Repeat steps 2 to 4 for 20 times, each time with a different generated network and thus a different baseline.

Step 6: Go back to step 1, update levels of the network-related parameters, and repeat steps 2 to 5.

Technical details: Overall, 192,000 instances of scenarios were simulated. All simulations were implemented in MATLAB *R*2019*b*. The computations were performed on a computer cluster with 27 computing nodes (all nodes connect to IBM Spectrum Scale with Infiniband), with a total of 720 cores, and approximately 7.4 TB of RAM. Total execution time was 355 hours.

5.2.1. Results

This section presents the main results of the experimental analysis. The impacts of the policy parameters, such as enforcing social distancing or expanding testing capacities, on each of the four measures used to describe the epidemic outcomes are estimated using ordinary least squares regression. The measurements of the

[24] The choice of baseline condition is arbitrary and does not affect the final results because only differences in the computed outputs are compared.

[25] For each time period t in the time window $t = 1, 2, \cdots, T$, construct $\mathcal{I}' \subseteq \mathcal{I}$ by excluding the recovered and perished cases from $\mathcal{I}$. Count the number of currently hospitalized individuals and update $\phi_{eff}(t)$ and $\rho_i(t)$ according to Eq. (9). For each pair of individuals $i, j \in \mathcal{I}'$ with $S_i(t) =$ susceptible and $S_j(t) =$ infected, generate the number of physical contacts between them at period t, if they are directly linked in the generated network, according to $N_{ij}(t) \sim Poisson(\lambda_{ij})$, $\lambda_{ij} = 1/d_{ij}^2$, and $d_{ij} = \frac{\alpha_i e^{n\alpha_i} + \alpha_j e^{n\alpha_j}}{e^{n\alpha_i} + e^{n\alpha_j} - \alpha_i e^{n\alpha_i} + \alpha_j e^{n\alpha_j}}$, which is the smooth approximation of the max function, and update the status of i according to $Bernoulli\big(1 - (1-\delta)^{N_{ij}(t)}\big)$. For an already infected individual j, $S_j(t) =$ infected, update the status to detected if the incubation period W_i has been reached and update the number of hospitalized individuals. For a detected individual j, update the status if time period H_i of hospitalization has been reached, and then assign that person to the recovered state with a probability $\rho_i(t)$ or to the perished state with a probability of $1 - \rho_i(t)$. Update $\mathcal{I}'$.

Table 3
Parameters and their levels in the experimental analysis ($|\mathcal{I}| = 1000$).

Network structure			Infection	Health status		Healthcare system			Social distancing		
γ	p	q	δ	(θ_0, θ_1)	η	ω	ϕ	κ	a_c	a_{nc}	π
{ 0.30, 0.70 }	{ 0.25, 0.50 }	$p/100$	0.20	{ (5, 5), (8, 2) }	0.50	{ 0.30, 0.76 }	{ 0.40, 0.80 }	{ 50, 100 }	5	5	{ 0.80, 0.95, 1 }

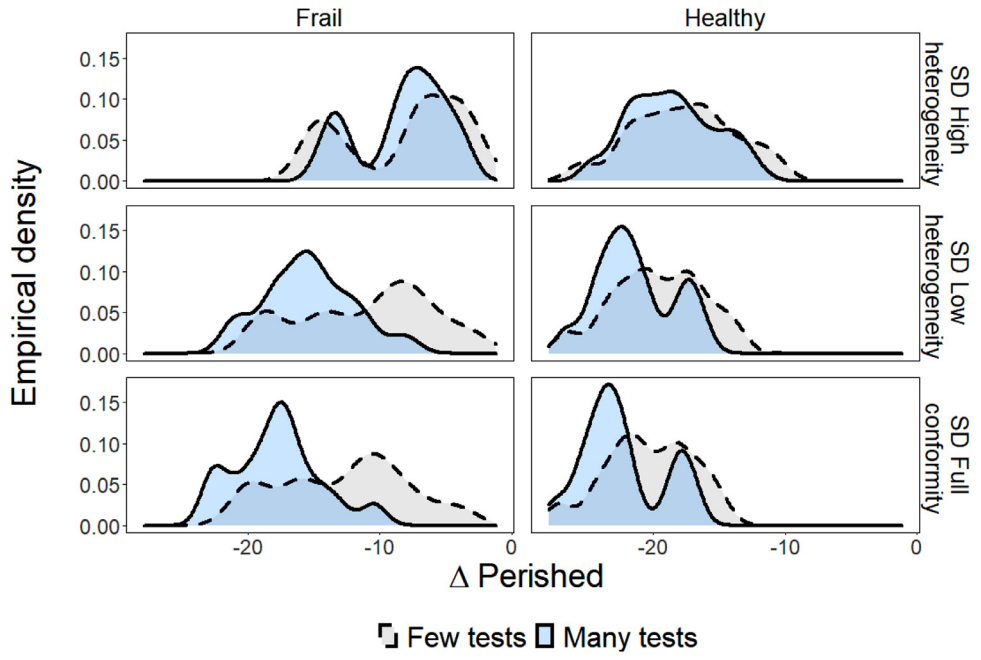

Fig. 5. The empirical densities of the reduction in the total number of perished cases (cutback in the number of perished cases with reference to the baseline condition) as a function of the heterogeneity in social distancing behavior (rows), population health status (columns), and testing capacities (ω) – gray filled dashed line: small testing capacity; blue filled solid line: large testing capacity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

outcomes all refer to the baseline condition, so the parameter estimates represent changes in the respective outputs obtained by adopting a health intervention method.

• *Saving Human Lives*

This analysis focuses on reducing the number of deaths. Fig. 5 exhibits the distribution of the improvements for different conditions, defined by population health status and heterogeneity in social distancing behavior. The empirical distributions suggest several insights.

First, the distributions of the many versus few tests conditions, depicted in blue and gray, indicate that effectiveness of expanding testing capacity depends on the extent of social distancing. When compliance with social distancing and self-isolation requirements (top row) are heterogeneous, the two distributions of testing capacity conditions almost entirely coincide. The distributions diverge as the level of heterogeneity decreases. Second, in the few tests condition specifically, no considerable changes across rows appear, regardless of variations in the social distancing conditions. That is, enforcing social distancing does not pay off if it is not combined with extensive testing capacities. Improvements in the distributions of the many tests are prominent with less heterogeneity in social distancing behavior. Third, these observations hold irrespective of the population's overall health status, according to the comparisons of the distributions across columns.

Enforcing social distancing is a practical, yet, costly strategy to contain the virus and reduce the number of infected cases. Nevertheless, considering the significant economic implications, determining precisely when this costly strategy is most effective may be relevant. In other words, do improvements in containing the virus always translate into perceptible improvements in the number of deaths?

As suggested by Fig. 6, in some situations, they do not. Specifically, among frail segment of the population, a decline in the number of infected agents directly reduces the number of deaths (high slopes in the left panel), but, unless the healthcare system is fully prepared for the crisis, in terms of capacity and care quality, this substitution rate appears relatively independent of the characteristics of the healthcare system (parallel lines in the left panel). [26] Among healthy segment of the population, a slight improvement in the number of infected agents reduces the number of deaths only marginally, and it makes no additional difference if improved further. These insights have direct policy implications. For a vulnerable population, the focus must be on reducing the number of infected agents, e.g., by enforcing self-isolation and implementing mobility restrictions. But if the population's health status is somewhat better, perhaps due to a partially effective vaccine, strategies that were effective for reducing the number of infected cases become less pressing (logarithmic patterns in the right-side panel), irrespective of the current state of the healthcare system (parallel curves in the right-side panel). In this situation, reducing deaths requires redirecting resources toward alternative strategies, such as enhancing healthcare quality (shifting the curves in the right-side panel upward).

[26] Data are aggregated across the four network types. Observations from different network types are not necessarily uniformly distributed along the horizontal axis. Specifically, a greater number of the data points from the centralized and dense networks are likely to appear for smaller versus larger values of ΔInfection.

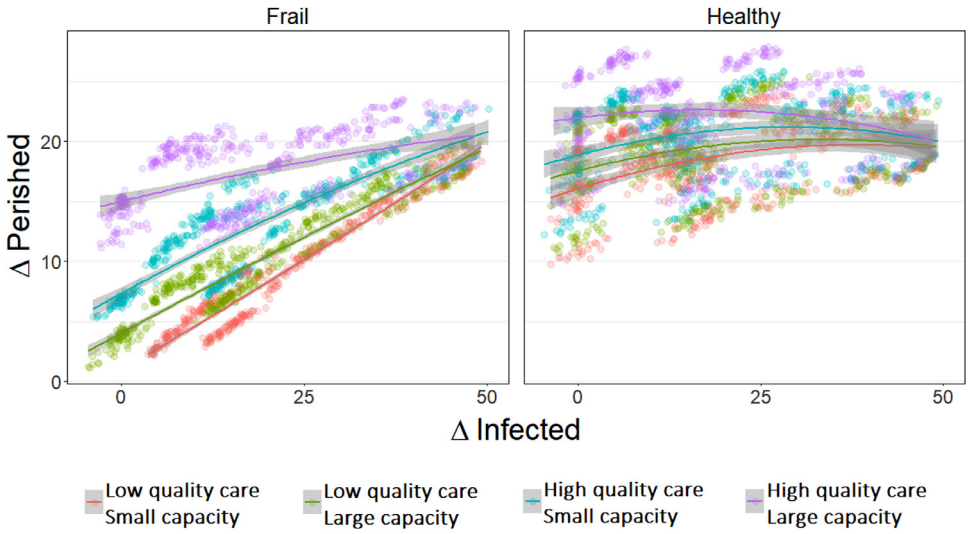

Fig. 6. (Color online) Improvements in death rates (population percentage) as a function of reducing infections (population percentage), for frail (left panel) and healthy (right panel) populations. Different lines correspond to different healthcare system characteristics. Shaded areas are the 99% confidence intervals..

• *Pushing the Peak to the Future*

An important policy objective, especially at an early stage of a pandemic, is delaying the outbreak's peak to provide more time to increase necessary capacities to handle the crisis without compromising human lives. An effective strategy would push the peak forward without stretching out the overall pandemic duration. Therefore, this analysis focuses on the recovery pace metric ν, and its determinants, while controlling for the epidemic duration.

By controlling for changes in pandemic duration, testing capacity, and network parameters, as Table Table 4 shows, the analysis demonstrates a perfect mediation relationship among social distancing, changes in the total number of infected agents over the entire pandemic period, and the recovery pace as an indicator of the relative time location of the peak. [27] Model 1 demonstrates the significance of social distancing heterogeneity for the recovery pace (p-value$<$ 0.0001; adjusted R^2 increases by 10% compared with Model 0 that includes only control variables). Model 2 shows that heterogeneity also informs predictions of *improvements* in the total number of infected agents over the entire pandemic period. The impact on recovery pace vanishes when such improvements are accounted for though, which implies a perfect mediation process. The total impact of (a) decreasing social distancing heterogeneity on (c) recovery pace (0.04 in Model 1) is the product of the impact of (a) on (b) improvements in the total number of infected agents, that is 21.68 in Model 2, and the impact of (b) on (c), which is 0.002 in Model 3 plus the controlled impact of (a) on (c) in Model 3, which in turn is not significant statistically, as Fig. 7 indicates. This is an instance of a classical perfect-mediation (Baron & Kenny, 1986).

Both key output differences, $\Delta\nu$ and ΔInfected, are measured simultaneously at the end of the period and determined endogenously by the system's overall behavior. Thus, the direction of causality in the right side of the Fig. 7 cannot be verified. Nevertheless, after controlling for the pandemic duration, the peak time is delayed by reducing the total number of infected agents. Having accounted for this effect, enforcing social distancing no longer has a direct impact. Furthermore, improving the testing capacities effectively pushes the peak time to the future, after controlling for the improvements in the total number of infected agents.

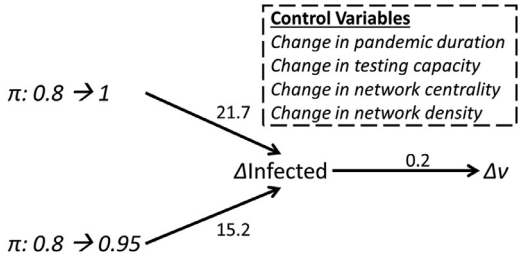

Fig. 7. Illustration of the mediation analysis in Table 4. Reducing heterogeneity in social distancing delays the pandemic peak by lowering the number of infected agents. Improving testing capacity remains the most impactful control variable.

• *Network Structure Matters*

In this section, the analysis addresses how the relationships detailed in the previous two sections might depend on the network structure, by analyzing output data for different settings defined by the network parameters.[28] Figs. 8 and 9 contain the results of a series of regression analyses for, respectively, reductions in the number of deaths and the number of infected cases, in terms of population percentages.[29] The results concerning death rates are separated across the four types of network structure, determined according to network centrality and community sparsity levels, and the population's overall health status. For the infected cases, similar patterns were observed for the healthy and frail populations, so observations from both these groups are combined for the parameter estimations. The same holds for the results of the epidemic duration, presented in Table 5. Therefore, the impact of intervention methods on epidemic duration is not contingent on the population's overall health status. For the recovery pace though, the only factor that introduces contingencies to the impact of intervention methods is network centrality, as Table 6 details.

[27] For the choice of the control variables, different nested regression models were explored and compared according to the adjusted R-squared criterion.

[28] All the independent variables in these series of regression models are categorical. Thus, the analyses are essentially analyses of variance.

[29] Each point in the graph represents a coefficient estimate. Thus, Fig. 8, for instance, summarizes results of eight regression models.

Table 4
Ordinary least square estimates of the four regression models. Reducing the total number of infected agents (ΔInfected) is a perfect mediator for the relationship between enforcing social distancing and delaying the peak time of a pandemic ($\Delta\nu$), controlling for other factors. All the estimated effects are significant at < 0.0001 except for the two bold estimates in Model 3 (p-value> 0.05).

	Dependent variable:							
	Model 0: $\Delta\nu$		Model 1: $\Delta\nu$		Model 2: ΔInfected		Model 3: $\Delta\nu$	
	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.
Main variables								
SD Full conformity (π: 0.8 → 1)			4.0	0.10	21.7	0.17	**-0.4**	0.22
SD Low heterogeneity (π: 0.8 → 0.95)			3.1	0.09	15.2	0.16	**0.0**	0.16
Δ Infected							0.2	0.01
Control variables								
Δ pandemic duration	0.9	0.02	0.6	0.01	-0.7	0.02	0.7	0.01
Many tests (ω: 0.3 → 0.76)	21.0	0.28	16.1	0.26	8.1	0.44	14.5	0.26
Centralized (γ: 0.3 → 0.7)	-2.3	0.09	-2.1	0.07	-7.9	0.12	-0.5	0.10
Dense (p: 0.25 → 0.50)	-1.5	0.09	-1.2	0.07	-3.8	0.13	-0.4	0.08
Intercept	3.8	0.09	2.0	0.09	4.3	0.15	1.1	0.09
Model Fit								
Adjusted R^2	69.2		78.9		92.8		81.4	
RMSE	2.68		2.22		3.7340		2.08	

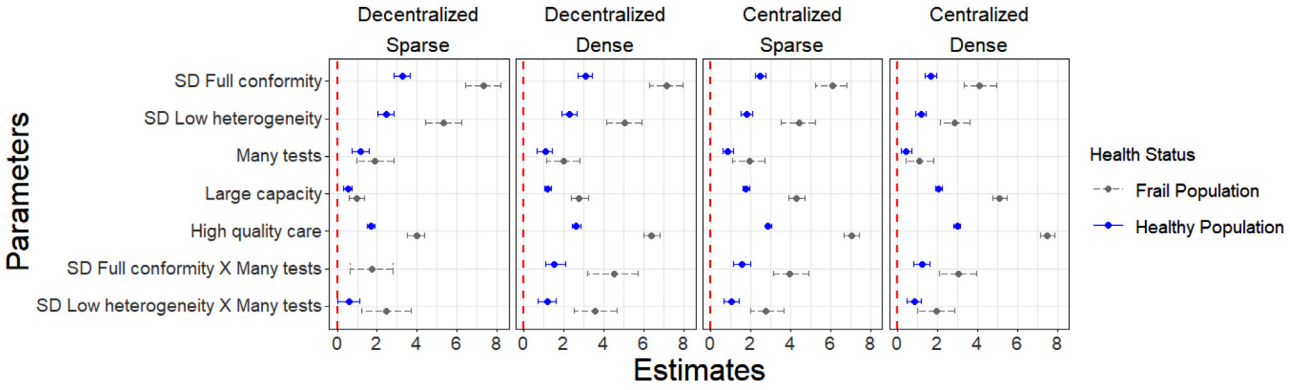

Fig. 8. Coefficient plots of ordinary least squares parameter estimates of the impact of intervention methods on total deaths relative to the baseline condition. Sample size $n = 460$ for each of the 4 regression models corresponding to the frail population, and $n = 480$ for each of the 4 regressions corresponding to the healthy population. The 99% confidence intervals are computed by bootstrapping and using the bias-corrected and accelerated percentile method with 1000 bootstrap samples. The R^2 values corresponding to decentralized networks are as follows: for dense, 92% and 90%, and for sparse, 87% and 82%, for the healthy and frail populations respectively. These values are $\sim 92\%$ for all four models corresponding to the centralized networks.

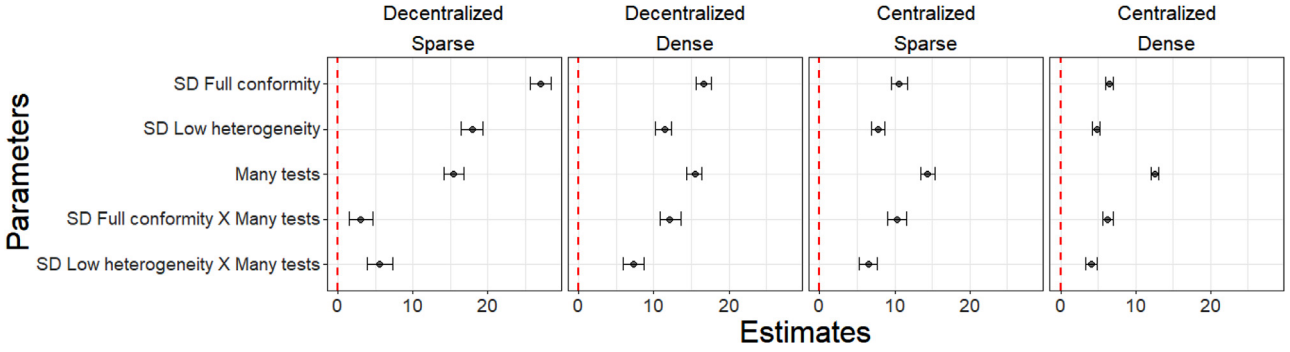

Fig. 9. Coefficient plots of ordinary least square parameter estimates of the impact of intervention methods on total infected cases relative to the baseline condition. Sample size $n = 940$ for all four models. The 99% confidence intervals are computed by bootstrapping and using the bias-corrected and accelerated percentile method with 1000 bootstrap samples. The R^2 values are greater than 98% for all models.

In general, a system with frail populations is more responsive to the health intervention methods in terms of the number of deaths. As network centrality and community density increase, the importance of healthcare quality and capacity increase too, whereas relative importance of social distancing decreases unless combined with an extensive testing capacity. This finding is particularly important given the difficulty and considerable economic and social consequences of enforcing social distancing in centralized and dense networks. In other words, this costly intervention method can be justified, in the presence of a high degree of network centrality and community density, only when an extensive testing capacity is also present. In line with the evidence provided by Fig. 5 and Table 4, testing capacity is a critical determinant of different outcomes. Furthermore, the joint effect of social distancing and many tests in centralized networks decreases by 25% if only 5% of the population stop complying with social distanc-

Table 5
Ordinary least square parameter estimates of the impact of intervention methods on changes in epidemic duration (change in T with reference to the baseline condition). Model $R^2 \geq 98\%$ and sample size $n = 940$ for all models.

	Decentralized						Centralized					
	Sparse			Dense			Sparse			Dense		
	Est.	99% Conf. Int.		Est.	99% Conf. Int.		Est.	99% Conf. Int.		Est	99% Conf. Int.	
SD Full conformity	6.5	5.1	7.9	7.6	6.7	8.5	6.9	6.2	7.7	3.8	3.2	4.4
SD Low heterogeneity	4.9	3.4	6.1	4.1	3.5	4.8	3.6	2.8	4.3	2.3	1.8	2.9
Many Tests	-15.6	-16.7	-14.5	-14.6	-15.2	-14.0	-14.2	-14.9	-13.5	-13.5	-14.0	-12.9
SD Full conformity × Many tests	-11.3	-13.0	-9.8	-5.0	-6.0	-4.0	-4.6	-5.5	-3.8	-0.8	-1.4	-0.1
SD Low heterogeneity × Many tests	-6.0	-7.4	-4.5	-2.1	-3.0	-1.3	-1.9	-2.7	-1.1	-0.8	-1.4	-0.2

Table 6
Ordinary least square parameter estimates of the impact of intervention methods on changes in recovery pace (change in ν with reference to the baseline condition). Model $R^2 = 93\%$ for centralized and 72% for decentralized setting, and sample size $n = 1880$ for both models.

	Decentralized			Centralized		
	Est.	99% Conf. Int.		Est.	99% Conf. Int.	
SD Full conformity	11	10.1	12.0	4	3.7	4.6
SD Low heterogeneity	8	7.1	8.6	3	3.0	3.6
Many Tests	9	8.4	9.8	8	7.2	7.9
SD Full conformity × Many tests	-9	-10.6	-8.1	2	1.6	3.0
SD Low heterogeneity × Many tests	-5	-6.1	-4.1	1	0.1	1.1

ing requirements. This observation holds for both healthy and frail populations, such that the results suggest the need to shift focus toward healthcare capacity and care quality when operating in a more centralized network and with denser communities. If social distancing measures are going to be adopted in such settings, they must be combined with expanding testing capacities and implemented with strict surveillance.

The results also demonstrate some direct and indirect benefits of investing in expanding testing capacity. First, the direct impact of increasing the number of tests is independent of the population's health status, regardless of intervention methods and across various outcome measures. Second, the direct impact of increasing the number of tests is relatively small for reducing the number of deaths, but it remains highly effective as an intervention method, in that it moderates the impact of social distancing. Third, more tests also help reduce the number of infections and increase the recovery pace, shifting the peak to the right side of the time scale. Forth, it has the strongest effect for reducing epidemic duration, irrespective of the network structure, such that it can mitigate the adverse economic consequences of a strict self-isolation policy.

6. Conclusions

An effective response to infectious disease outbreaks demands the inclusion of features of the embedding environment. The characteristics of the population, their overall health status, heterogeneity in compliance with new regulations, patterns of contacts, and heterogeneity in other individual-level attributes interact with policy parameters to yield the ultimate outcomes. In turn, policy measures alter features of the environment. For example, enforcing compliance with social distancing likely changes patterns of contacts. Because there is no one-size-fits-all solution for all the variations in environments that exist in reality, any effective strategy should be designed with careful consideration of interrelationships among key factors.

The unified framework presented herein integrates critical components of a contagion process with the key characteristics of the embedding environment and accounts for heterogeneity in individual-level attributes and sources of uncertainty at different layers of the system. The model is primarily intended to inform policy decisions concerning public health interventions. To this objective, the model construction process explicitly aimed to integrate the main features of the contagion model with the characteristics of the environment and relevant policy parameters. Components of the model are subsequently synthesized in a broader ABM that can account for heterogeneity in these micro-level attributes that collectively yield the macro-level outcomes.

Among contextual factors, patterns of contacts among individuals, as established in the network structure, require particular consideration. These patterns directly control virus transmission throughout the population, and as the experiment results of this study reveal, the network structure interacts with policy parameters, such that their effectiveness is strongly contingent on the characteristics of the network. In the proposed model, the network structure is characterized by two simple factors: the degree of centrality in the network topology and the density of connections within components of the network, called communities. With these two factors, it is possible to generate random networks using an algorithm that is introduced here. The empirical results, which exhibit reasonably narrow confidence intervals when centrality and density are fixed, show that these two factors capture a large proportion of the variation in the outcome stemming from the inherent randomness in the network. The results demonstrate the critical role of the network structure as a substantial element for designing an effective response to infectious disease outbreaks.

Beyond these methodological contributions, a series of stylized examples illustrate the applicability of the model. An extensive controlled experiment offers insights into the overall tendencies of various types of outcomes and the relative effectiveness of different policy parameters according to contextual factors. The results highlight the importance of social distancing and self-isolation. When network centrality and density increase, the impact of these intervention methods is increasingly contingent on the presence of extensive testing capacity. Moreover, expanding testing capacity is essential to mitigate the impact of social distancing on epidemic duration, which subsequently might translate into economic losses. Finally, in terms of the number of deaths, when network centrality and density increase, healthcare system capacity and care quality dominate other factors; social distancing prevails in decentralized and sparse networks.

The proposed framework might be extended in various directions, such as by incorporating macro-level covariates to explain heterogeneity in the individual-level attributes. Income distribution, population pyramid type, access or usage patterns of high-speed internet, lifestyle parameters, and dimensions of national culture all offer viable candidates for investigation. Continued research also might aim to reduce the computational demands of the model, which tends to be a significant limitation of the practical use of ABMs as a whole. Recent criticisms of the ABM paradigm tend to highlight its calibration and validation issues, which can restrict the scope of its findings to a qualitative and suggestive level (Leombruni & Richiardi, 2005; Marks, 2007). To address these concerns, a growing body of research deals with estimation techniques based on simulated minimum distance (Fabretti, 2013), which attempt to minimize some distance between the observed and simulated data (Marks, 2013). They usually entail a considerable computational cost. In the context of the model developed in this paper, it would be interesting to exploit the method of simulated moments (Banerjee, Chandrasekhar, Duflo, & Jackson, 2013; McFadden, 1989) and explore the potential of nonparametric simulated maximum likelihood combined with Bayesian inference. Due to the computational benefits of such approaches, it may become possible to exploit the informational content of the data more fully (Grazzini, Richiardi, & Tsionas, 2017; Kukacka & Barunik, 2017). Finally, the computational experiment, similar to other ABMs, provides qualitative insights to inform policy decisions; it does not mandate any specific decision. Compromises among policy decision consequences, including those quantified here but also cost and implementation convenience, as well as the benefits of potential enrichment with additional context-specific and time-location idiosyncratic considerations, might be revealed and clarified by social multi-criteria evaluations Munda (2004) and the multiple-criteria decision-aiding paradigm (Ghaderi, Ruiz, & Agell, 2017; Greco, Figueira, & Ehrgott, 2016; Kadziński, Ghaderi, & Dabrowski, 2020; Roy, 1990).

Acknowledgments

I am grateful to three anonymous reviewers for their helpful comments. I acknowledge financial support from the Spanish Agencia Estatal de Investigación (AEI), through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S).

References

Abbe, E. (2017). Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research, 18*(1), 6446–6531.
Abrahamson, E., & Rosenkopf, L. (1997). Social network effects on the extent of innovation diffusion: A computer simulation. *Organization science, 8*(3), 289–309.
Albert, R., Jeong, H., & Barabási, A.-L. (1999). Diameter of the world-wide web. *Nature, 401*(6749), 130–131.
Aleman, D. M., Wibisono, T. G., & Schwartz, B. (2011). A nonhomogeneous agent-based simulation approach to modeling the spread of disease in a pandemic outbreak. *Interfaces, 41*(3), 301–315.
Atkeson, A. (2020). What will be the economic impact of COVID-19 in the US? Rough estimates of disease scenarios. *Technical Report*. National Bureau of Economic Research.
Auchincloss, A. H., & Diez Roux, A. V. (2008). A new tool for epidemiology: The usefulness of dynamic-agent models in understanding place effects on health. *American journal of epidemiology, 168*(1), 1–8.
Axelrod, R. (2006). Agent-based modeling as a bridge between disciplines. *Handbook of computational economics, 2*, 1565–1584.
Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). The diffusion of microfinance. *Science (New York, N.Y.), 341*(6144).
Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science (New York, N.Y.), 286*(5439), 509–512.
Barabási, A.-L., Albert, R., & Jeong, H. (2000). Scale-free characteristics of random networks: The topology of the world-wide web. *Physica A: Statistical Mechanics and its Applications, 281*(1–4), 69–77.
Barabási, A.-L., & Bonabeau, E. (2003). Scale-free networks. *Scientific American, 288*(5), 60–69.
Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations.. *Journal of personality and social psychology, 51*(6), 1173.
Berger, D. W., Herkenhoff, K. F., & Mongey, S. (2020). An seir infectious disease model with testing and conditional quarantine. *Technical Report*. National Bureau of Economic Research.
Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the national academy of sciences, 99*(suppl 3), 7280–7287.
Broido, A. D., & Clauset, A. (2019). Scale-free networks are rare. *Nature communications, 10*(1), 1–10.
Van den Bulte, C., & Stremersch, S. (2004). Social contagion and income heterogeneity in new product diffusion: A meta-analytic test. *Marketing Science, 23*(4), 530–544.
Burt, R. S. (1980). Models of network structure. *Annual review of sociology, 6*(1), 79–141.
Chiou, L., & Tucker, C. (2020). Social distancing, internet access and inequality. *Technical Report*. National Bureau of Economic Research.
Davis, G. F. (1991). Agents without principles? the spread of the poison pill through the intercorporate network. *Administrative science quarterly*, 583–613.
Easley, D., Kleinberg, J., et al. (2010). *Networks, crowds, and markets*: 8. Cambridge university press Cambridge.
El-Sayed, A. M., Scarborough, P., Seemann, L., & Galea, S. (2012). Social network analysis and agent-based modeling in social epidemiology. *Epidemiologic Perspectives & Innovations, 9*(1), 1.
Epstein, J. M. (2006). *Generative social science: Studies in agent-based computational modeling*. Princeton University Press.
Epstein, J. M. (2009). Modelling to contain pandemics. *Nature, 460*(7256), 687.
Erdõs, P., & Rényi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci, 5*(1), 17–60.
Fabretti, A. (2013). On the problem of calibrating an agent based model for financial markets. *Journal of Economic Interaction and Coordination, 8*(2), 277–293.
Fernandes, N. (2020). Economic effects of coronavirus outbreak (covid-19) on the world economy. In *Available at SSRN 3557504*.
Fornaro, L., & Wolf, M. (2020). *Covid-19 coronavirus and macroeconomic policy*.
Fristedt, B. (1993). The structure of random partitions of large integers. *Transactions of the American Mathematical Ssociety, 337*(2), 703–735.
Ghaderi, M., Ruiz, F., & Agell, N. (2017). A linear programming approach for learning non-monotonic additive value functions in multiple criteria decision aiding. *European Journal of Operational Research, 259*(3), 1073–1084.
Gilbert, E. N. (1959). Random graphs. *The Annals of Mathematical Statistics, 30*(4), 1141–1144.
Goldenberg, A., Zheng, A. X., Fienberg, S. E., Airoldi, E. M., et al. (2010). A survey of statistical network models. *Foundations and Trends® in Machine Learning, 2*(2), 129–233.
Goldenberg, J., Han, S., Lehmann, D. R., & Hong, J. W. (2009). The role of hubs in the adoption process. *Journal of Marketing, 73*(2), 1–13.
Grazzini, J., Richiardi, M. G., & Tsionas, M. (2017). Bayesian estimation of agent-based models. *Journal of Economic Dynamics and Control, 77*, 26–47.
Greco, S., Figueira, J., & Ehrgott, M. (2016). *Multiple criteria decision analysis*: 37. Springer.
Haber, M. J., Shay, D. K., Davis, X. M., Patel, R., Jin, X., Weintraub, E., … Thompson, W. W. (2007). Effectiveness of interventions to reduce contact rates during a simulated influenza pandemic. *Emerging infectious diseases, 13*(4), 581.
Helbing, D., Brockmann, D., Chadefaux, T., Donnay, K., Blanke, U., Woolley-Meza, O., … Schutte, S., et al. (2015). Saving human lives: What complexity science and information systems can contribute. *Journal of Statistical Physics, 158*(3), 735–781.
Helgeson, C., Srikrishnan, V., Keller, K., & Tuana, N. (2020). *Why simpler computer simulation models can be epistemically better for informing decisions*.
Hellmann, T., & Staudigl, M. (2014). Evolution of social networks. *European Journal of Operational Research, 234*(3), 583–596.
Hethcote, H. W. (2000). The mathematics of infectious diseases. *SIAM Review, 42*(4), 599–653.
Ip, E. H., Rahmandad, H., Shoham, D. A., Hammond, R., Huang, T. T.-K., Wang, Y., & Mabry, P. L. (2013). Reconciling statistical and systems science approaches to public health. *Health Education & Behavior, 40*(1_suppl), 123S–131S.
Iyengar, R., Van den Bulte, C., & Valente, T. W. (2011). Opinion leadership and social contagion in new product diffusion. *Marketing Science, 30*(2), 195–212.
Jackson, M. O. (2010). *Social and economic networks*. Princeton university press.
Jackson, M. O. (2019). *The human network: How your social position determines your power, beliefs, and behaviors*. Pantheon.
Kadziński, M., Ghaderi, M., & Dabrowski, M. (2020). Contingent preference disaggregation model for multiple criteria sorting problem. *European Journal of Operational Research, 281*(2), 369–387.
Karrer, B., & Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical review E, 83*(1), 016107.
Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. *Proceedings of the ninth acm sigkdd international conference on knowledge discovery and data mining* (pp. 137–146).
Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character, 115*(772), 700–721.
Kukacka, J., & Barunik, J. (2017). Estimation of financial agent-based models with simulated maximum likelihood. *Journal of Economic Dynamics and Control, 85*, 21–45.

Landler, M., & Castle, S. (2020). Behind the virus report that jarred the us and the uk to action. *The New York Times, 2*.
Lee, J.-S., Filatova, T., Ligmann-Zielinska, A., Hassani-Mahmooei, B., Stonedahl, F., Lorscheid, I., … Parker, D. C. (2015). The complexities of agent-based modeling output analysis. *Journal of Artificial Societies and Social Simulation, 18*(4).
Leombruni, R., & Richiardi, M. (2005). Why are economists sceptical about agent-based simulations? *Physica A: Statistical Mechanics and its Applications, 355*(1), 103–109.
Macal, C. M. (2016). Everything you need to know about agent-based modelling and simulation. *Journal of Simulation, 10*(2), 144–156.
Mahajan, V. (2010). Innovation diffusion. *Wiley International Encyclopedia of Marketing*.
Manshadi, V., Misra, S., & Rodilitz, S. (2020). Diffusion in random networks: Impact of degree distribution. *Operations Research*. Doi: 10.1287/opre.2019.1945
Marks, R. E. (2007). Validating simulation models: A general framework and four applied examples. *Computational Economics, 30*(3), 265–290.
Marks, R. E. (2013). Validation and model selection: Three similarity measures compared. *Complexity Economics, 2*(1), 41–61.
Marshall, B. D., & Galea, S. (2015). Formalizing the role of agent-based modeling in causal inference and epidemiology. *American Journal of Epidemiology, 181*(2), 92–99.
McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society*, 995–1026.
McKibbin, W. J., & Fernando, R. (2020). *The global macroeconomic impacts of covid-19: Seven scenarios*.
Miller, J. H., Page, S. E., & Page, S. (2009). *Complex adaptive systems*. Princeton university press.
Muller, E., & Peres, R. (2019). The effect of social networks structure on innovation performance: A review and directions for research. *International Journal of Research in Marketing, 36*(1), 3–19.
Munda, G. (2004). Social multi-criteria evaluation: Methodological foundations and operational consequences. *European Journal of Operational Research, 158*(3), 662–677.
Newman, M. (2003a). Random graphs as models of networks. *Handbook of Graphs and Networks, 1*, 35–68.
Newman, M. (2003b). The structure and function of complex networks. *SIAM Review, 45*(2), 167–256.
Newman, M. (2018). *Networks*. Oxford University Press.
Obiols-Homs, F. (2020). Precaution, Social Distancing and Tests in a Model of Epidemic Disease. *Technical Report*.
Platt, D. (2020). A comparison of economic agent-based model calibration methods. *Journal of Economic Dynamics and Control, 113*, 103859.
Rahmandad, H., & Sterman, J. (2008). Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Management Science, 54*(5), 998–1014.
Riley, S. (2007). Large-scale spatial-transmission models of infectious disease. *Science (New York, N.Y.), 316*(5829), 1298–1301.
Roy, B. (1990). Decision-aid and decision-making. *European Journal of Operational Research, 45*(2–3), 324–331.
Sculli, D., & Wong, K. (1985). The maximum and sum of two beta variables and the analysis of pert networks. *Omega, 13*(3), 233–240.
Tracy, M., Cerdá, M., & Keyes, K. M. (2018). Agent-based modeling in public health: Current applications and future directions. *Annual Review of Public Health, 39*, 77–94.
Utomo, D. S., Onggo, B. S., & Eldridge, S. (2018). Applications of agent-based modelling and simulation in the agri-food supply chains. *European journal of operational research, 269*(3), 794–805.
Vallverdú, J. (2014). What are simulations? an epistemological approach. *Procedia Technology, 13*, 6–15.
Willinger, W., Alderson, D., & Doyle, J. C. (2009). Mathematics and the internet: A source of enormous confusion and great potential. *Notices of the American Mathematical Society, 56*(5), 586–599.
Zheng, H., Son, Y.-J., Chiu, Y.-C., Head, L., Feng, Y., Xi, H., … Hickman, M., et al. (2013). *A primer for agent-based simulation and modeling in transportation applications*.